Using CMIP6 temperature emulators for optimal climate policy under model uncertainty

Hannah Römer[*], Yu Huang[†], Raphael Römer [‡]

October 2, 2025

The substantial spread in climate projections from large-scale climate models complicates the derivation of optimal climate policies, but is also an important reminder of the limited predictability of the climate system. We build Integrated Assessment Models (IAMs) reflecting this range of temperature projections using novel linear response theory and Lasso regression-based CMIP6 temperature emulators. This allows a quantification of the policy impact of concern over climate model uncertainty under a range of preference specifications. Under expected utility maximization, the optimal policy includes full decarbonization from 2075 onwards, even though this is only the case for 30% of CMIP6 models when modeled separately. Mitigation increases with robustness concerns, but decreases when weighting climate models by historical performance.

1. Introduction

Optimal climate policy is based on the trade-off between the costs and benefits of emission reduction and consequently has at its core a physical model of the climate response to anthropogenic emissions. However, detailed climate models, so-called general circulation models (GCMs), disagree significantly with each other, an important observation from the Coupled Model Intercomparison Project Phase 6 (CMIP6). The complexity of the climate system, the difficulties of modeling highly non-linear multi-dimensional processes, and finally the limited observations of global climate indicators all contribute to the persistence of this disagreement, so that it seems unreasonable to expect a convergence of model projections in the foreseeable future.

[*]Department of Economics, University of Oxford
[†]Earth Systems Modelling Group, Technical University of Munich
[‡]Department of Mathematics and Statistics, University of Exeter

Taking as given the limited predictability of the climate system, the question emerges how much policy making should account for the uncertainty over different climate models. In this paper, we take a pragmatic approach by asking how much climate policy depends on how a planner accounts for model uncertainty. Leveraging novel temperature emulators for a range of CMIP6 models, we can assess the quantitative implications for the optimal emission mitigation path of several margins in the planner's objective function. Our contributions are two-fold: First, we introduce machine learning based temperature response functions for a range of 31 different models included in CMIP6. Second, we couple the entire set of these approximations with an economic module, which allows us to obtain optimal policy paths not only for each temperature model separately, but also under uncertainty over temperature models. A natural comparison is the one between the choices of a planner who accounts for the uncertainty explicitly to one who chooses the mean between model projections. We examine two additional channels: first, whether the planner exhibits an explicit preference for robust decision making, and second, whether the planner's subjective probability of a specific climate model is based on its historical performance.

We find that our climate model approximations turn out to be highly accurate across a range of different test cases. Evaluations of the temperature responses are consistent with the tests suggested by Folini et al. (2024). Coupling the economic module with each temperature model separately leads to a large spread in optimal policy paths. For example, the year after which full mitigation is optimal ranges from 2050 to 2120. The mean over these implied policy paths deviates from the choice of an expected utility maximizing planner who puts equal weights on each model, in particular in the long-run, with the latter choosing full mitigation starting in 2075. In particular, explicitly accounting for model uncertainty in an optimization problem with concave utility has a large effect on the policy choice. The more a planner cares about robustness, the more they avoid policies with adverse outcomes in any state of the world, so that the optimal mitigation rate increases. If the planner assigns subjective probabilities to climate models based on their historical performance, both the variance and the mean of temperature projections decreases, pushing down mitigation. We uncover a strong interplay between robustness concern and a high variance in subjective weights.

Our climate model emulation approach is based on linear response theory and imposes as the only assumption that the dynamics of the global mean surface temperature $T(t)$ approximately solve a system of the form

$$LT(t) = F(t)$$

where L is a linear differential operator and F is time dependent radiative forcing caused by greenhouse gases. Under this assumption, the temperature $T(t)$ can be computed as a convolution between the time-series of radiative forcing and the so-called *Green's function*. Approximately, the temperature response is thus a linear function in all previous levels of

radiative forcing. This is a slightly less restrictive assumption than the often used box-model, which implicitly requires the Green's Function to be a sum of exponentials. While previous studies (e.g. Aengenheyster et al. (2018), Smith et al. (2018)) relied on experiments with abrupt CO_2 doubling or annual increases in CO_2 by 1% to calibrate the Green's function, our machine learning based approach also admits other scenarios to be used in training the model to learn the response function coefficients.

Importantly for the economic application, the resulting approximations are reduced enough to be included in an optimization procedure, so that it is possible to solve simultaneously for the optimal emissions and consumption path. However, as the temperature response relies on the entire time series of emissions, the planner's problem includes many lags of emissions as state variables. Due to the curse of dimensionality, it is thus not feasible to solve the problem recursively. We argue in section 4.2 that it is reasonable to assume no learning about the correct climate model during the time of interest and abstract from re-optimizing.

It is not obvious (and neither provable nor falsifiable) that the range of CMIP6 models actually covers the true range of possible climate outcomes. Nevertheless, they are our best estimates for plausible scenarios and thus the best feasible basis for policy making. The uncertainty over the correct climate model is, importantly, a matter of Knightian uncertainty, not risk. In particular, there is no way to attribute an objective probability to each model. We implement and compare two different choices for prior subjective probability distributions: An agnostic who thinks all models are equally likely, and an empiricist who puts more weight on those models that have performed well with regard to historical observations. Allowing for robustness concerns in the objective function further accounts for the idea that a planner is more averse to deep uncertainty than risk.

Internal variability and model spread Usually, a natural choice to sort out wrong models would be to check whether they are capable of reproducing measured data well. With a model for the global mean surface temperature, this approach turns out to be more complicated as the climate shows internal variability as discussed in Ghil and Lucarini (2020) and Dijkstra (2013), i.e. infinitely many other climate trajectories might have happened given the same historic emissions. Often, this can be linked to chaotic dynamics as pointed out in von der Heydt et al. (2021) and Houghton et al. (2001). This implies that even if a climate model did not reproduce the past global mean surface temperature (GMT) well, it is not necessarily worse than models reproducing the measured GMT time series more accurately. Instead, a model that does not reproduce past observations but is rooted in sensible physical theory might contain information on which other states the climate could have been in and potentially will be in in the future. Thus, we argue that the internal variability of the climate justifies using models that have not performed well historically, and we are convinced that it is important to take the spread of CMIP6 models seriously to some extent to make well-informed policy

decisions.

Relevant literature The economics literature has long emphasized the relevance of uncertainty in climate projections for sensible climate policy making. Cai and Lontzek (2019) include uncertain tipping points as well as long-run risk in their extension of the DICE model, building on the work by Nordhaus (2008). Distinguishing the uncertainty surrounding climate change from risk has been promoted by Brock and Hansen (2018). They advocate for a further distinction between ambiguity, as the situation where probabilities across models are unknown and misspecification, to refer to unknown flaws within the models themselves. They also develop a theoretical framework to include the specific kind of uncertainty posed by climate change in a decision problem, building on the smooth ambiguity aversion in recursive preferences by Klibanoff et al. (2009) as well as the robustness literature, see Hansen and Sargent (2011). Similarly, Millner et al. (2012) push for deviating from expected utility theory to appropriately account for the uncertainty posed by climate change. Complementary to our study is the work by Zhao et al. (2023), Rudik (2020) and Barnett et al. (2020) who solve recursive optimization problems allowing for Knightian uncertainty over a small number of parameters. Our approximation approach imposes less stringent assumptions on temperature emulators, but prevents a recursive solution method.

Structure of the paper The structure of the paper is as follows. Section 2 explains the estimation of CMIP6 temperature emulators as well as the model we use to relate emissions to carbon concentrations. We demonstrate the accuracy of our approximations using a variety of test cases in section 3. Section 4 introduces our economic model, based on the DICE model by Nordhaus (2008), and discusses the objective function and resulting decision problem of the social planner. Figure 1 illustrates the full model pipeline. The implications of different model specifications on policy variables are shown in section 5. Section 6 concludes.

2. Climate Component: CMIP6 temperature emulators

The climate component of the fully coupled integrated assessment model consists of the bottom line shown in figure 1. The core of the climate component is the linear response theory-based model that predicts a temperature time series from a time series of radiative forcing. This part is built to capture as many of the intricate dynamics as possible of the CMIP6 climate models.

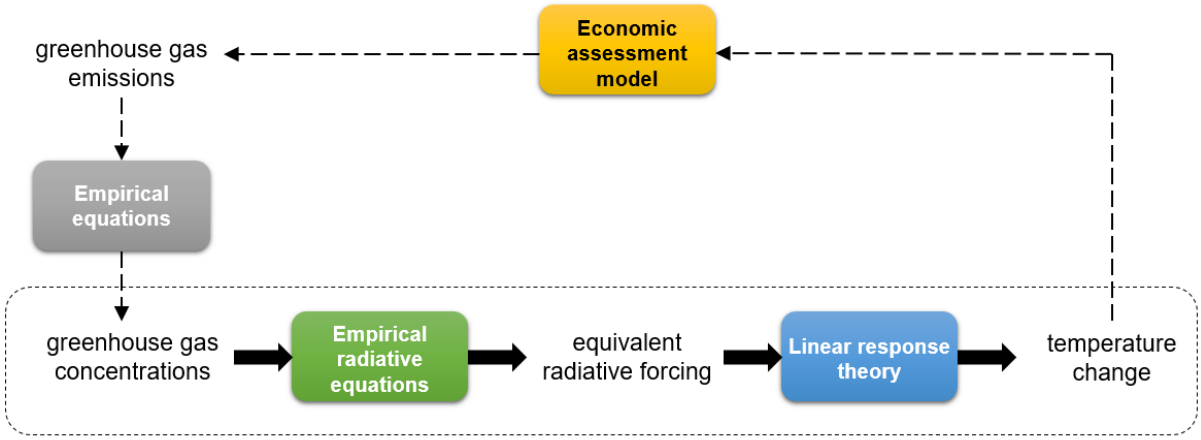


Figure 1: Pipeline of the full model

2.1. Estimating the Temperature Response to Radiative Forcing

2.1.1. Linear Response Theory

Linear response theory generally describes how a system of the form $LT(t) = F(t)$ with L being a linear differential operator responds to a time-dependent forcing $F(t)$. We assume that the dynamics of the global mean surface temperature GMT (here $T(t)$) in CMIP6 models can be approximated by such an equation. $F(t)$ is the forcing by greenhouse gases that causes $T(t)$ to change, and the underlying CMIP6 temperature dynamics are approximated by the action of some linear differential operator L. This modeling assumption worked well for CMIP5 models (Aengenheyster et al. (2018)), and our results suggest the same for CMIP6 models. Solving the above equation leads to the following expression for the temperature change $\Delta T(t)$ as a result of a time-dependent forcing $F(t)$:

$$\Delta T(t) = \int_0^t G(t')\Delta F(t - t')dt'$$

It is a convolution of the forcing F with the Green's function G, which describes the GMT's present-day response to all previous levels of radiative forcing and thus contains the information about memory effects.

2.1.2. Training Phase for Coefficients Estimation

The response theory model can be written in a discrete form as:

$$\Delta T(t) = \sum_{n=0}^{t} g_n \Delta F_n$$

where $n = 0, ..., t$ denote the discrete times (in unit of year) until the time t, then ΔF_n is the time series of the radiative forcing anomaly (compared to the pre-industrial radiative forcing), $\Delta T(t)$ is the temperature anomaly (compared to the pre-industrial temperature) as the prediction target, and g_n represents the Green's function coefficients to be determined ahead.

To make use of this model to predict the temperature, one needs to first estimate the model's coefficients by fitting it based on the existing data (hereafter called training phase). The temperature data corresponding to different SSP scenarios are for each considered model acquired from the CMIP6 project. Correspondingly, the data for ΔF_n is available from http://greenhousegases.science.unimelb.edu.au (based on Meinshausen et al. (2017) and Meinshausen et al. (2020)), which aligns with the design for the CMIP6 future scenarios experiments and the shared socio-economic pathway (SSP) greenhouse gas concentrations. For each specific CMIP6 climate model, the time series of ΔF and ΔT are paired within each set of SSP scenarios. 31 climate models from CMIP6 are picked, and each of their data is used to estimate the coefficients of a response theory model. Hence, we obtain a set of 31 response theory models, each based on one underlying CMIP6 model.

During the training phase, the data from two CMIP6 greenhouse gas experiments (that is, abrupt-4xCO2 and SSP-126) were used to estimate the model coefficients within a Lasso regression. The abrupt-4xCO2 experiment is commonly used for estimating response theory model coefficients. We additionally incorporate the SSP-126 data, which includes CO2 removal.

2.1.3. Evaluation and Testing Phases

After the coefficients of a response theory model were determined, the data from three other CMIP6 experiments, i.e., 1pctCO2, SSP-245 and SSP-585, which were unseen during the training phase, were used to evaluate the model accuracy (hereafter called evaluation phase). The response theory models can make consistent temperature predictions for the CO2 increase scenario of 1pctCO2, across the dataset of the 31 employed CMIP6 models. For the SSP-245 and SSP-585 data, when comparing the temperature time series predicted by the response theory models with the ground truth (i.e., the native CMIP6 dataset), there is an overall additive shift in the absolute value of the predicted time series, but the increasing trends (the first derivatives) during 2014-2100 are consistent between the predicted and ground truth temperature time series. We calculate the difference in the mean value between the predicted and ground truth temperature time series, and find it to be identical for the evaluation dataset of SSP-245 and SSP-585. This difference is attributed to a cooling bias in CMIP6 models whose causes are a subject of ongoing research (see e.g., Zhang et al. (2021), Xu et al. (2021), and Flynn et al. (2023)) and is partially attributed to the effects of aerosol forcing during 1850-2014

- in addition to the greenhouse gas forcing - unconsidered in the response theory models.

As we find this difference between the SSP scenarios in the RT models and the CMIP6 models to be constant for different SSP scenarios, it can be bias-corrected by subtracting this constant from the predicted temperature time series across all SSP scenarios and added to the predicted temperature time series for the abrupt 4xCO2 experiment. We test the response theory model using the dataset for the SSP-370 scenario (testing phase), and find that the predicted temperature time series for 2014-2100 is consistent with the ground truth. It is noted that the CMIP6 SSP experiments for 2014-2100 are driven by greenhouse gas forcing without consideration of aerosol forcing, which explains the consistent trends between CMIP6's native time series and our response theory model's predictions.

2.2. Radiative Forcing and non-CO2 greenhouse gases

Our response theory model takes radiative forcing as an input which can be derived from greenhouse gas concentrations in the atmosphere. The carbon concentration in the atmosphere is given endogenously within our economic model. For the most relevant other greenhouse gases, CH4, N2O, CFCs and HFCs, we follow the projections of the SSP-370 scenario. For each featured greenhouse gas, the corresponding radiative forcing can be calculated following Etminan et al. (2016).

2.3. The carbon cycle model

Just as in DICE, the output of the economic model is carbon dioxide emissions, so that the carbon cycle has to be modeled separately. All CO_2 emissions up to time t are summarized by the vector $\mathbf{E}_t = (E_{-T_0}, \ldots, E_t)$ where $-T_0$ is the start of industrialization (which we take to be the year 1850). The atmospheric carbon concentration at time t can then be written as

$$M_t = g(\mathbf{E}_t) \tag{1}$$

for some function g. We follow Joos et al. (2013) in partitioning the system into a finite number of boxes. The functional form of g is then assumed to be given by

$$M_t = M_0 + \sum_{s+0}^{t} \sum_{i=0}^{N} \psi_i \lambda_i^{t-s} E_s$$

where M_0 is the initial, pre-industrial amount of carbon in the atmosphere. This specification accounts for the different time scales how long emissions stay in the atmosphere.

	Box 1	Box 2	Box 3	Box 4
ψ	0.22	0.22	0.28	0.28
$\log(0.5)/\log(\lambda_i)$		277	25	3

Table 1: Parameter values used for carbon cycle model

2.3.1. Carbon cycle calibration

The calibration of the carbon cycle follows the best fit CMIP5 ensemble considered in Dietz et al. (2021). There are $N = 4$ boxes. The first box is the permanent one, meaning that $\lambda_1 = 1$ and ψ_1 is the fraction of emissions permanently staying in the atmosphere. For all $i > 1$, the time it takes for half of the carbon to leave box i is given by $\log(0.5)/\log(\lambda_i)$. The parameters for the carbon cycle are collected in table 1.

3. Performance of the Climate Component

To evaluate the performance of our RT-based temperature emulators, we assess their performance in the test cases outlined in Folini et al. (2024). They developed a comprehensive framework to test reduced climate models for economic assessments while accounting for the low signal-to-noise ratio prominent in climate data, which prevents a straight-forward evaluation based on existing historical data. The tests are thus largely based on computational experiments as the ones used in our estimation of the model. This section expands on each of the test cases separately.

Note that the suggested tests in Folini et al. (2024) are benchmarks in comparison to CMIP5 models, whereas we use CMIP6 data here. As outlined in Folini et al. (2024), there should not be a major difference between the models in CMIP5 or CMIP6. We note, however, that some of the CMIP6 models produce higher temperatures than their previous versions in CMIP5, given the same forcing. This is a topic of ongoing research and, for example, Tokarska et al. (2020) proposes that observational data can be used to constrain the range of CMIP6 models to be consistent with CMIP5. As we explain below in section 4.2, we specifically decide to also include CMIP6 models that are not strongly supported by historical data, including those with significantly higher temperatures than their CMIP5 versions.

3.1. Carbon cycle calibration: temporal evolution of instantaneous release of 100 GtC

The first suggested test targets the calibration of the carbon cycle to match the benchmark data from Joos et al. (2013) on the temporal evolution of an instantaneous release of 100Gt carbon to the present-day atmosphere. As our carbon cycle model is directly based on Joos

et al. (2013), this test is trivially satisfied by our calibration.

3.2. Temperature emulator calibration: warming in response to quadrupling of CO2

The second test is aimed at assessing the emulator's temperature in response to a sudden quadrupling of the CO2 concentration. This should match that of the CMIP6 models under the same forcing. The data from the abrupt-4xCO2 experiment in CMIP6 was directly used in training the RT-model and is matched well. This is shown for the CMIP6 model "IPSL" in Figure 2.

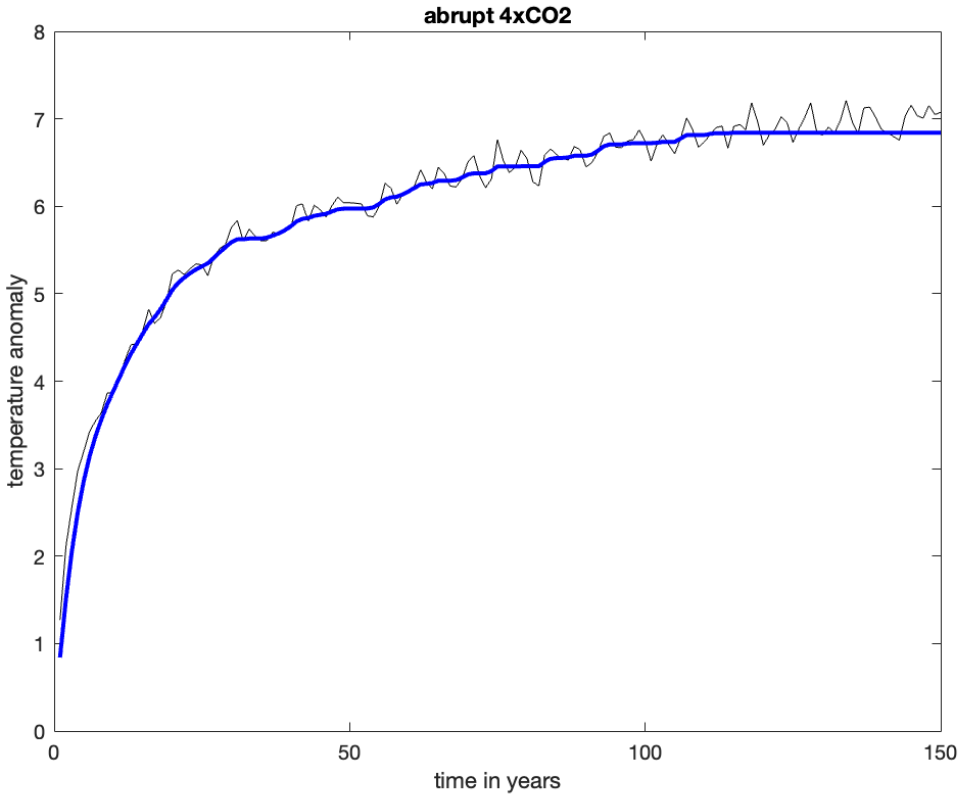


Figure 2: Temperatures of the CMIP6 model "IPSL" and the corresponding RT model in response to an abrupt quadrupling of the CO2 concentration in the atmosphere.

3.3. Transient climate response: 1pctCO2 experiment

The third test is checking whether the temperature increase of the RT models as a result of a yearly increase of 1% in CO_2 concentration, i.e., the transient climate response, reproduces that of the CMIP6 models. Figure 3 shows that it is captured well, except for the CMIP6 model CMCC-ESM2. *To be fixed.*

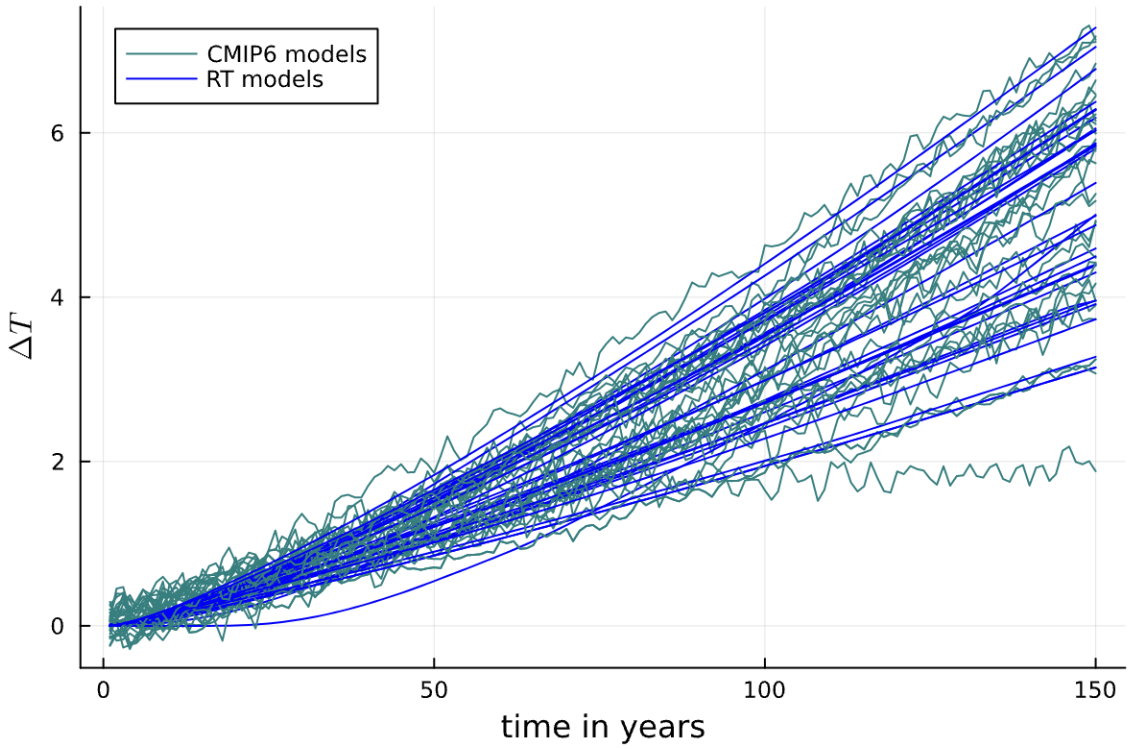


Figure 3: CMIP6 model temperatures and the corresponding RT model temperatures in response to a yearly increase of 1% in CO_2 emissions.

3.4. SSP scenarios

The fourth and final test suggested by Folini et al. (2024) checks how the RT models perform in relation to the historical period and future SSP scenarios. Figure 4 shows good agreement of the RT model for the CMIP6 model "IPSL".

3.5. Concluding remarks on the performance of the climate component

We conclude that the RT models reproduce the temperature dynamics of the considered CMIP6 models well. Still, some of the CMIP6 models themselves do not perform well on the checks with respect to the CMIP5 suggested by Folini et al. (2024). As we noted in section 3.2, this is subject of ongoing research, and we maintain that all models should still be considered in order to make well informed policy decisions as outlined in the introduction.

4. Integrated climate economics model

An integrated assessment model is made up of an economic model which describes output, emissions and damages from temperature increases together with a climate model which maps these emissions into temperatures. Imposing the objective to maximize some predefined wel-

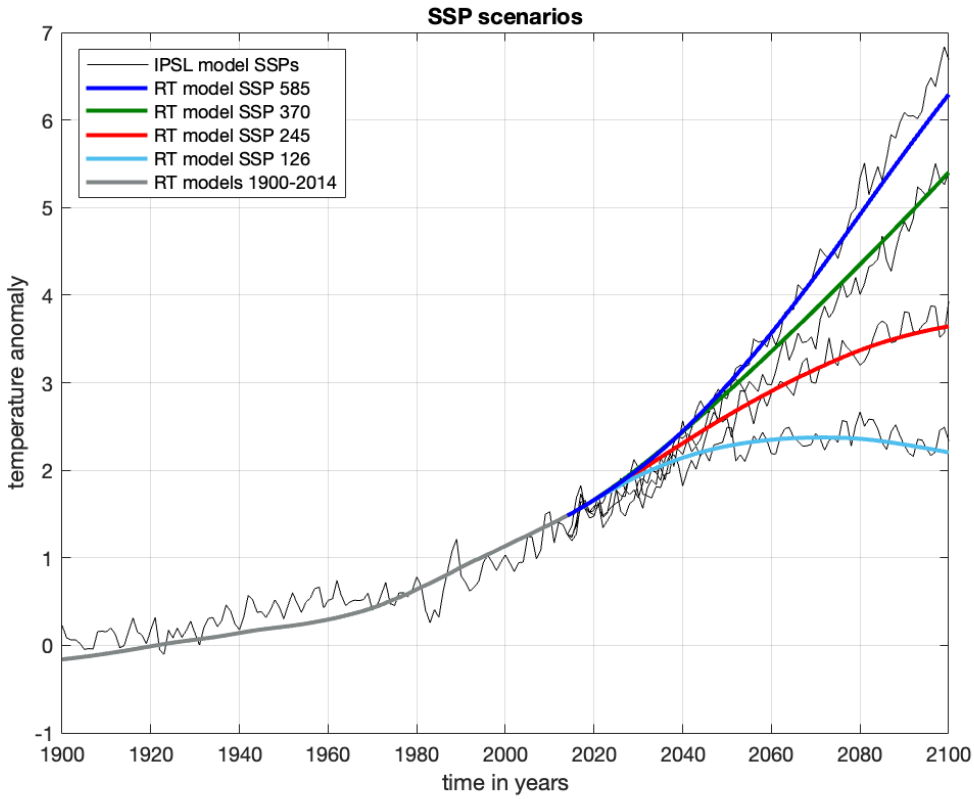


Figure 4: RT model reproducing the trends of the CMIP6 model "IPSL" (black) for four different SSP scenarios.

fare function enables us to derive optimal trajectories for mitigation and consumption.

We incorporate the temperature emulators derived above into the DICE model developed by Nordhaus (2008).

4.1. The economic model

Output is produced by combining capital K and labor L in a Cobb-Douglas fashion with a possibly time-varying productivity parameter A_t:

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}. \tag{2}$$

Production processes generate emissions. We only endogenize carbon dioxide emissions and take other greenhouse gases as given, see section 2. In line with the DICE model, the relationship between CO_2 emissions and production is assumed to be linear with a time dependent parameter σ_t. The variable $0 \leq \mu_t \leq 1$ captures the share of emissions that is mitigated.

Overall emissions are then given by

$$E_t = \sigma_t(1-\mu_t)AK_t^{\alpha}L_t^{1-\alpha} + E_t^{EX} \tag{3}$$

where E_t^{EX} are exogenous submissions not related to economic activity. Mitigation comes at a cost, measured as a fraction of output:

$$\Lambda(\mu_t) = m_t\mu_t^{\overline{m}}, \tag{4}$$

where m_t and $\overline{m}$ are exogeneous parameters.

Finally, climate damages are assumed to follow a quadratic function

$$D_t = \pi_2 T_t^2. \tag{5}$$

To be explicit about the dependence on a specific CMIP6 model θ used in the computation of temperature, we sometimes write T_t^{θ}. Note that the modeled temperature, even in 2020, may differ substantially from observed data, see figure 6. Damage estimates are based on the temperatures of the underlying climate RT model.

Output after both mitigation expenses and damages is thus given by

$$Y_t = (1-D_t)(1-\Lambda(\mu_t))AK_t^{\alpha}L_t^{1-\alpha}.$$

4.2. The decision problem of the social planner

A social planner chooses savings and mitigation time paths $(\mathbf{s}, \mu)$ to optimize current welfare given some utility function U. Under any fixed temperature model θ, the planner's optimization problem is given by

$$\begin{aligned}
&\max_{s_t,\ \mu_t} \sum_{t=0}^{\infty} \beta^t U(C_t) \\
&\text{s.t. } C_t = (1-s_t)Y_t && (6)\\
&\quad K_{t+1} = s_t Y_t + (1-\delta)K_t && (7)\\
&\quad Y_t = (1-D_t(T_t))(1-\Lambda(\mu_t))AK_t^{\alpha}L_t^{1-\alpha} && (8)\\
&\quad E_t^{IND} = \sigma_t(1-\mu_t)AK_t^{\alpha}L_t^{1-\alpha},\ E_t = E_t^{IND} + E^{LAND} && (9)\\
&\quad M_T = g(\mathbf{E}_t),\ T_t = f_{\theta}(\mathbf{M}_t) && (10)\\
&\quad 0 \le \mu_t \le 1, \quad 0 \le s_t \le 1. && (11)
\end{aligned}$$

For any feasible savings and mitigation path $(\mathbf{s}, \mu)$ and a fixed model θ, we define the associated present value of social welfare as

$$V(\mathbf{s},\ \mu,\ \theta) = \sum_{t=0}^{\infty} \beta^t U(C_t).$$

4.2.1. Objective function under subjective expectations

A social planner who accounts for different models θ instead solves an optimization problem of the form

$$\max_{s,\mu} \mathbb{E}\left\{V(s,\ \mu,\ \theta)\right\}. \tag{12}$$

An important determinant of the optimal allocation is the distribution underlying the expectation in equation (12). However, there is no generally agreed upon objective probability distribution over the different models. Motivated by the theory of subjective expected utility and following previous work by Millner et al. (2012), we impose subjective probabilities as the basis for expectation formation in two ways:

(i) Agnostic expectations: All climate models are considered as being the true model with equal subjective probability.

(ii) Empiricist's expectations: The climate models are evaluated based on their historical performance. Those that match observations well are allocated a higher probability. The exact procedure to obtain these weights is documented in section 4.3.1.

Instead of advocating for one approach over the other, we contrast the two approaches here, in order to appreciate the discourse over the relevance of historical accuracy for climate models. While Tokarska et al. (2020) and Brunner et al. (2020) point out that weighting models by independence or performance in reproducing past warming trends can be used to constrain or alter future temperature projections, Bloch-Johnson et al. (2022) argues for the value of including also unexpectedly "hot" models, as long as physically correct, in order to account for possibly highly relevant climate scenarios.

We do not assume that the social planner changes decisions in the future and consequently do not formulate this recursively as a Bellman problem. Instead, the social planner makes a single decision today about the policy path and commits to it. This is partly due to computational concerns: The path dependence that was introduced by the convolution of the forcing with the Green's function leads to too many state variables which makes the problem computationally not feasible. Furthermore, many of the models we consider lead to trajectories that reach net-zero fast. Figure 5 reports the share of models under which our optimal policies include a full mitigation rate. As we do not consider CO_2 removal, one cannot lower emissions any further. Thus, new knowledge about the models after the point of full mitigation is irrelevant.

As the climate is defined as the moving average over 20 or 30 years (see for example [MET office current temperature]), temperature changes below this timescale are not particularly instructive for understanding a model's ability to capture the future climate. Conclusively, observed temperature in the near future is unlikely to give us new insights about the quality of the models as it is not clear which temperature changes are caused by internal variability, and which are an effect of radiative forcing. Without new information, updating does not make sense.

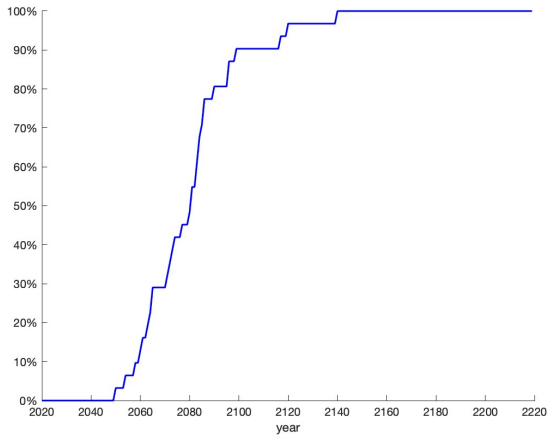


Figure 5: Share of models under which full mitigation is optimal

Note that we also assume that the social planner does not include today's observed temperature into their decision problem. This is consistent with the climate being a chaotic system, so that the current temperature being at a certain level does not mean that the climate will continue to follow a model that predicted this level for the current period. In the case of agnostic expectations, the social planner would also not choose to update their beliefs. They would only do so when weighting models according to their historical performance.

Finally, the planner also does not re-optimize based on realized output. Even when assuming that they would not learn from observing the exact damages, in principle, their choice in the future would still depend on the current economic state. Instead, we consider the choices of a planner who believes that they fully commit to a path a priori. Importantly, the policy choices are savings and mitigation **rates**, not quantities, so that they are feasible no matter the economic conditions.

4.2.2. Optimization under misspecification aversion

In an extension, we allow the social planner to be averse against model misspecification. The importance of robust decision making for climate economics has been emphasized by Brock and Hansen (2018).

The social planner assumes a prior distribution π^* over all climate models $\theta \in \Theta$. This prior distribution will either be uniform, which we call the agnostic prior, or based on the models' accuracy in matching historical temperature observations, an empiricist's prior. The planner also considers other distributions π, but penalizes deviation from the prior distribution. They then aim to find the policy such that the most adverse outcome **under this policy** is maximized. For tractability and in line with the literature, we assume that the penalization function is given by the relative entropy, or Kullback-Leibler divergence, between the distributions π and π^*. Formally, the objective becomes

$$\max_{s,\ \mu} \left\{ \min_{\pi} \left[\sum_{\theta} \pi(\theta) V(s,\ \mu,\ \theta) + \kappa \sum_{\theta} \pi(\theta) \log \left(\frac{\pi(\theta)}{\pi^*(\theta)} \right) \right] \right\}.$$

The specific functional form allows a closed form derivation of the minimizing distribution for any policy choice $(s,\ \mu)$ which is attained for

$$\pi(\theta) \propto \exp\left(-\frac{1}{\kappa} V(s,\ \mu,\ \theta) \right) \pi^*(\theta). \tag{13}$$

The probabilities are thus tilted towards the models with more adverse outcomes under the policy, i.e. those for which $V(s,\ \mu,\ \theta)$ is small. The maximization problem becomes

$$\max_{s,\ \mu} -\kappa \left\{ \log \left(\sum_{\theta} \pi^*(\theta) \exp\left(-\frac{1}{\kappa} V(s,\ \mu,\ \theta) \right) \right) \right\}.$$

The extent of the reweighting according to equation 13 is governed by κ. Small values of κ correspond to high concern over adverse outcomes. To the best of our knowledge, there has so far been no microfounded estimation of this crucial parameter[1]. We thus follow the parameter range in comparable studies, for example Rudik (2020) and Zhao et al. (2023), and consider $\kappa \in \{5,\ 10,\ 15\}$. Two special cases stand out:

- $\kappa \to 0$: There is no confidence in the prior, so that deviations are not penalized. The social planner prefers to take the action that maximizes the worst possible outcome. This is equivalent to the maxmin decision rule, see Gilboa and Schmeidler (1989).

- $\kappa \to \infty$: There is infinite high confidence in the prior, and thus infinite cost to deviate from it. The optimal choice thus coincides with the one under subjective expected utility and the prior distribution.

[1]Furthermore, it is not clear whether the value of κ should be pinned down by normative arguments or inferred from behavior.

4.2.3. The social cost of carbon

A key variable of interest is the social cost of carbon (SCC) which is the welfare cost of additional carbon dioxide emissions with a carbon content of 1 ton[2] today. This is given by the marginal rate of substitution between emissions and consumption, and thus can be written as the ratio between the two Lagrange multipliers on the emission constraint (3) and the budget constraint[3].

$$SCC_t = -\frac{\xi_t}{\lambda_t} = -\sum_{j=0}^{\infty} \beta^j \frac{U'(C_{t+j})}{U'(C_t)} \frac{\partial D_{t+j}}{\partial E_t} (1 - \Lambda(\mu_t)) A K_t^{\alpha} L_t^{1-\alpha}. \tag{14}$$

Under uncertainty, equation 14 has to also include the expectation operator. Under robustness concerns, the expectation is formed under subjective beliefs. Derivations can be found in appendix A.2.

4.3. Calibration

We follow Folini et al. (2024) in the calibration for the economic model who re-calibrated the Nordhaus (2008) model for a one year time step. Their model begins to run in 2015 while ours is recalibrated to start in 2020.
The labor process evolves according to

$$L_t = L_0 \exp(-\delta^L t) + L_\infty (1 - \exp(-\delta^L t)). \tag{15}$$

The productivity process is given by

$$A_t = A_0 \exp\left(\frac{g_0^A (1 - \exp(-\delta^A t))}{\delta^A} \right). \tag{16}$$

Note that to recalibrate the process to have $t = 0$ correspond with 2020, we use

$$A_{2020} = A_{2015} \exp\left(\frac{g_{2015}^A (1 - \exp(-5\delta^A))}{\delta^A} \right) \tag{17}$$

and

$$g_{2020}^A = \frac{\frac{dA_t}{dt}}{A_t}\Big|_{t=2020} = g_{2015}^A (1 - \exp(-5\delta^A)). \tag{18}$$

[2]The term social cost of carbon is used somewhat inconsistently in the literature, and sometimes refers to the social cost of carbon dioxide, as carbon is not emitted by itself. To obtain the social cost of carbon dioxide, our calculations have to be divided by 3.667. Carbon has a molecular weight of 12 g/mol, carbon dioxide has a molecular weight of 44 g/mol, so that the conversion factor is $44/12 \approx 3.667$.

[3]Note that splitting the budget constraint in two with the savings rate s implies two Lagrange multipliers which coincide at the optimum.

Parameter	Symbol	Value
Annual Rate of convergence (labor)	δ^L	0.0268
World population at 2020 (millions)	L_0	7917
Asymptotic world population (millions)	L_∞	11500
Growth rate for TFP per year (2020)	g_0^A	0.0005358
Decline of TFP growth per year	δ^A	0.005
Growth of carbon intensity per year (2020)	g_0^σ	-0.0153
Decline rate of decarbonization per year	δ^σ	0.001
Initial carbon intensity	σ_0	$8.855 \cdot 10^{-5}$

Table 2: Parameter values, exogenous processes

The carbon intensity declines over time and is given by

$$\sigma_t = \sigma_0 \exp\left(\frac{g_0^\sigma((1+\delta^\sigma)^t - 1)}{\delta^\sigma}\right). \tag{19}$$

Again, the updated values follow from

$$\sigma_{2020} = \sigma_{2015} \exp\left(\frac{g_{2015}^\sigma((1+\delta^\sigma)^5 - 1)}{\delta^\sigma}\right) \tag{20}$$

and

$$g_{2020}^\sigma = \frac{\frac{d\sigma_t}{dt}}{\sigma_t}\Big|_{t=2020} = g_{2015}^\sigma(1+\delta^\sigma)^5. \tag{21}$$

The backstop cost is given by

$$\theta_t = \frac{p_0^{back}(1+\exp(-g^{back}t))\sigma_t}{\overline{\theta}}. \tag{22}$$

Finally, we include land emissions which evolve according to

$$E_t^{LAND} = E_0^{LAND} \exp(-\delta^{LAND}t).$$

All relevant parameter values for the exogenous processes are collected in table 2.

For the damage function, we follow Haensel et al. (2020) in using the updated damage estimates from the meta study by Howard and Sterner (2017). This higher value includes a mark-up accounting for non-market damages. For comparison, we also report results for the damage function used by Folini et al. (2024) in the appendix. All other parameters needed for solving the economic model are collected in table 3.

Parameter	Symbol	Value
Capital depreciation rate (annually)	δ^K	0.1
Capital share	α	0.3
Exponent of control cost function	$\overline{\theta}$	2.6
Elasticity of intertemporal substitution	ρ	0.015
Discount factor	β	0.9852
Damage function coefficient	π_2	0.0074

Table 3: Parameter values

4.3.1. Weighting the climate models by their past performance

The weights for the empiricist's prior are pinned down by the climate models' past performance in matching aggregate temperature. We consider the aggregated squared residuals between the measured temperatures $T_{meas}(t)$ and the temperature $T_\theta(t)$ given by the RT-model θ:

$$r_\theta = \sum_{t=1900}^{2014} \Big(T_{meas}(t) - T_\theta(t)\Big)^2. \tag{23}$$

Then, we define normalized weights w_θ as

$$w_\theta = \frac{r_\theta^{-2}}{\sum_\theta r_\theta^{-2}}. \tag{24}$$

For each year t, one can then calculate a weighted mean temperature as

$$\bar{T}_{weighted}(t) = \sum_\theta w_\theta T_\theta(t). \tag{25}$$

Figure 6 shows the CMIP6 historic temperature anomaly, the measured temperature anomaly, and the RT model temperatures with the color intensity showing the weight of each model.

4.4. Computation

The economic model is solved using MATLAB's `fmincon` command and the default interior point algorithm, imposing the constraint that the choice variables have to lie in the unit interval. As an initial guess, we set $s_t = 0.25$ and μ_t as linearly increasing from 0.3 to 1 over the first 50 years and constantly 1 after. We solve an optimization problem for the first 200 years, so from 2020 to 2219. We then set the mitigation and savings rate to their respective value in 2219 and assume that these choices stay constant for another 250 years to obtain a terminal value function.

In the baseline, the algorithm converges for most temperature models. In the case of no

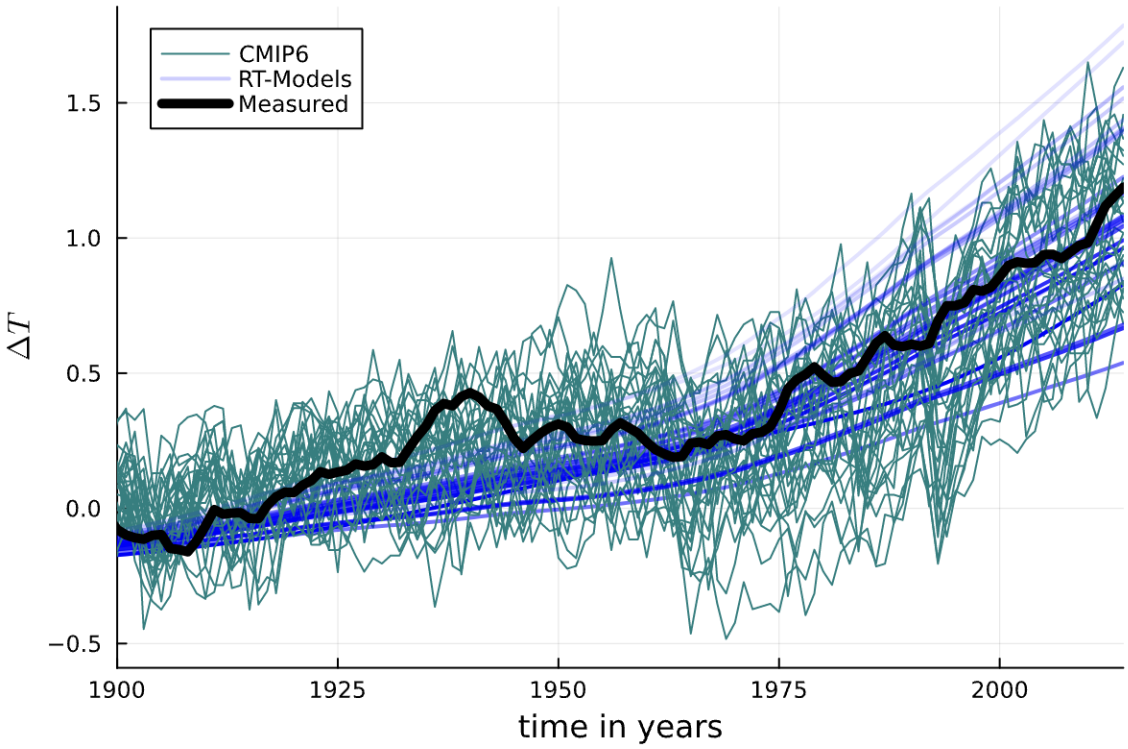


Figure 6: CMIP6 historic temperature anomaly (teal), the measured temperature anomaly (black), and the RT model temperatures (blue) over time in years. The intensity of the blue color of the plotted RT models indicates the strength of the calculated weights. The weaker the color the lower is the weight of this model.

convergence, the algorithm is reinitialized using the solution of a temperature model that predicts a similar path under the baseline SSP-370 scenario[4].

The derivative $\frac{\partial D_{t+j}}{\partial E_t}$ needed for the social cost of carbon is non-linear and has to be approximated. We can write

$$\frac{\partial D_{t+j}}{\partial E_t} = \frac{\partial D_{t+j}}{\partial T_{t+j}} \sum_{i=0}^{j} \frac{\partial T_{t+j}}{\partial E_t}. \tag{26}$$

and solve this using forward differences[5]. For a small h, we calculate $T^h_{t,t+j}$ as the temperature implied in $t+j$ from h more emissions in period t. The derivative is then approximated by

$$\frac{\partial T_{t+j}}{\partial E_t} \approx \frac{T^h_{t,t+j} - T_{t+j}}{h}.$$

We set $h = 10^{-5}$.

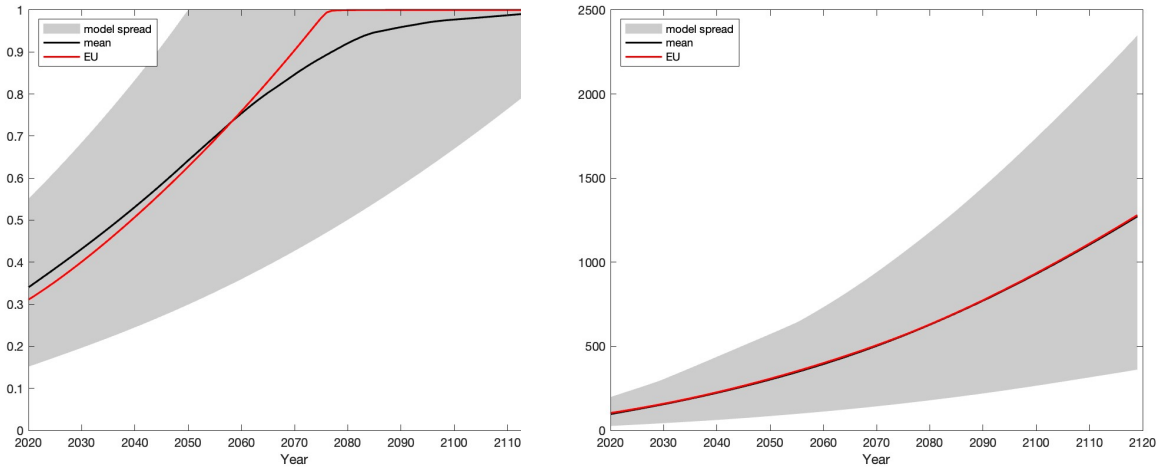
[4]The basic approach converges for all models in the case of Howard and Sterner (2017) damages, and for 27 out of 31 in the case of standard DICE damages.

[5]Note that it is sufficient here to find a one-sided derivative, as temperature is differentiable with respect to emissions.

5. Results from the Integrated Climate Economics Model

5.1. The model spread and accounting for uncertainty

In this section, we compare the spread of outcomes of a deterministic model assuming one of our CMIP6 approximations as a fixed temperature model with the outcomes of a social planner who is aware of the spread in models and makes optimal decisions under expected utility. The mitigation rates and SCCs are presented in figures 7a and 7b respectively. The gray bands indicate the spread of 31 different deterministic models. The black line is the mean of those trajectories, whereas the red line describes the path under expected utility maximization. Even though the mean of deterministic paths and the optimal mitigation rate are not far apart in the first half of the next century, we see a stronger divergence in later years. While there are models which even in the year 2100 do not recommend full mitigation, the expected utility social planner wants to mitigate fully starting in 2075. Some models advise to mitigate fully already by 2050.



(a) Mitigation Rate

(b) Social Cost of Carbon

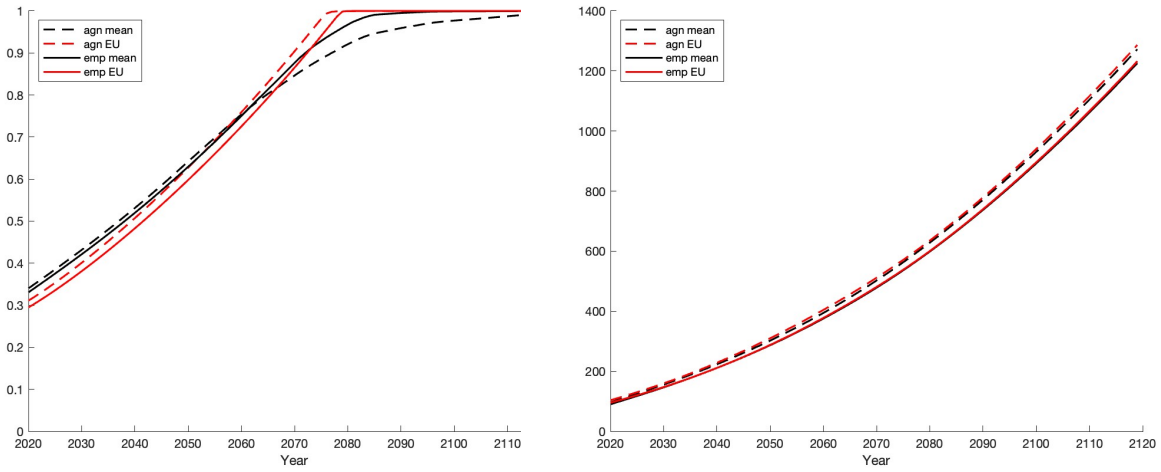
Figure 7: Results under agnostic prior

Considering the social cost of carbon, the differences between the mean of all trajectories implied from a single model and the trajectory of the optimal model is small but always negative. Due to risk aversion, the SCC is higher under expected utility than under the ex-post mean.

5.2. The implications of different weights

Figure 8 compares the implied optimal paths of the mitigation rate and the Social Cost of Carbon under the agnostic and empiricist's weights. Again, we report both the choices of a

social planner who solves an expected utility problem as well as the (appropriately weighted) mean of the paths under each model θ. Interestingly, under expected utility, the agnostic weighting (dashed line) will always lead to a higher optimal mitigation rate, which is not the case for the (weighted) means. Under agnostic weighting, the models which suggest very high warming are weighted more. The paths for the SCC are close and the agnostic preferences only lead to slightly higher values.



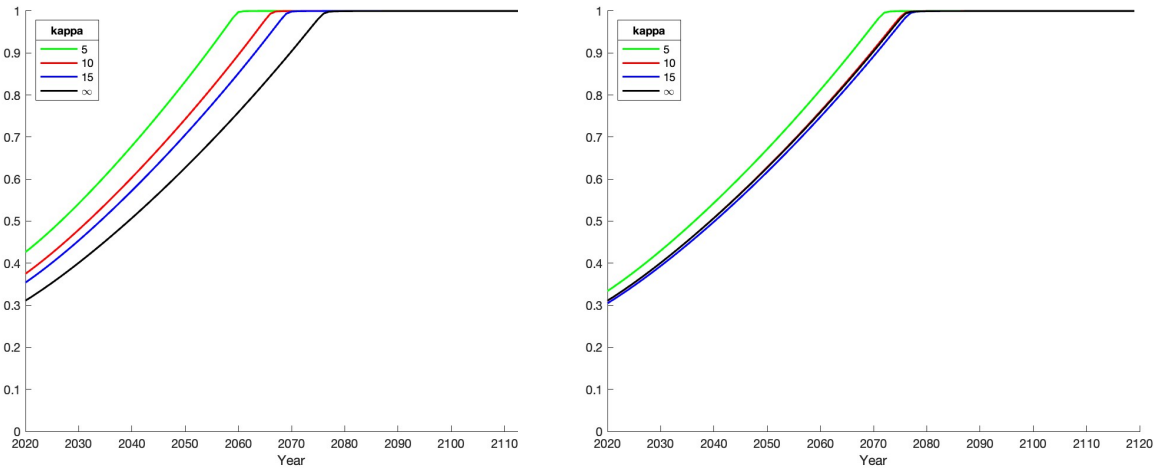
(a) Mitigation Rate (b) Social Cost of Carbon

Figure 8: Comparison between agnostic and empiricist's prior

5.3. The implications of robust preferences

A social planner who cares for robustness weighs the more adverse scenarios higher. In the present model, this implies that the models which project higher temperature increases are weighted more. Accordingly, this concern leads to a higher mitigation rate, and the mitigation rate becomes even higher for low levels of κ. Figure 9 shows the implied mitigation trajectories for both types of prior. In the case of an agnostic prior, all of them are substantially higher than the expected utility choice which is equivalent to a value of $\kappa = \infty$. With the lowest value for the robustness preference parameter $\kappa = 5$, the optimal mitigation rate in our first period 2020 is already more than 10% higher than in the baseline. Furthermore, full mitigation is already optimal in the year 2060. The effect of robust preferences is significantly dampened in the case of weights rooted in historical performance of the models. Only for the lowest value of κ, there is a strong effect which again leads to a higher mitigation rate and earlier full mitigation. Remember that robust preferences reweigh the different models based on how adverse the consequences under a certain policy are and how much weight was given to them under the prior distribution. The models which lead to the most adverse outcomes under the expected utility maximizing policy paths (i.e., those with the hottest projected

temperature) are weighted less under the empiricist prior, so that they are less relevant under robust preferences.



(a) agnostic prior (b) empiricist's prior

Figure 9: Share of mitigation under different robustness preference parameters

6. Conclusion

In this paper, we incorporate state-of-the-art approximations of global climate models into the widely-used DICE model. We can then examine both the spread in optimal economic policies under each of these climate models, and the decision of a policy maker who is aware of this uncertainty. A crucial aspect of any decision under uncertainty is a sensible choice of weights and objective functions. We present and compare an agnostic and empiricist approach in picking weights as well as robustness concerns in the objective function.

As the climate system continues to exhibit deep uncertainty, and may even include chaotic elements, a convergence in temperature projections from different GCMs seems neither feasible nor necessarily desirable. It is thus imperative to consider the effect of uncertainty over climate models for policy, an undertaking that has so far been hampered by computational constraints. The ease and tractability of our approximation method however allows the derivation of optimal policies under all available knowledge, which includes the disagreement between climate models.

Furthermore, we can derive the implications of model uncertainty under a range of objective functions, without taking a normative stance on how much a planner *should* care about it. For example, our work illustrates the interaction between robust preferences and the choice of prior weights: mitigation policy varies little both with a change in priors under expected utility maximization, and with a change in the robustness parameter under the empiricist's

prior. There is however significant variation when combining robust decision making with the agnostic prior, with both a higher mean and variance in projected temperature compared to the empiricist's weighting.

References

Aengenheyster, M., Feng, Q. Y., van der Ploeg, F., and Dijkstra, H. A. (2018). The point of no return for climate action: effects of climate uncertainty and risk tolerance. *Earth System Dynamics*, 9(3):1085–1095.

Barnett, M., Brock, W., and Hansen, L. P. (2020). Pricing uncertainty induced by climate change. *The Review of Financial Studies*, 33(3):1024–1066.

Bloch-Johnson, J., Rugenstein, M., Gregory, J., Cael, B. B., and Andrews, T. (2022). Climate impact assessments should not discount 'hot' models. *Nature*, 608(7924):667–667. Bandiera_abtest: a Cg_type: Correspondence Publisher: Nature Publishing Group Subject_term: Climate sciences, Climate change.

Brock, W. A. and Hansen, L. P. (2018). Wrestling with uncertainty in climate economic models. *Becker Friedman Institute for Economics Working Paper No. 2019-71*.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R. (2020). Reduced global warming from cmip6 projections when weighting models by performance and independence. *Earth System Dynamics*, 11(4):995–1012.

Cai, Y. and Lontzek, T. S. (2019). The social cost of carbon with economic and climate risks. *Journal of Political Economy*, 127:2684–2734.

Dietz, S., van der Ploeg, F., Rezai, A., and Venmans, F. (2021). Are economists getting climate dynamics right and does it matter? *Journal of the Association of Environmental and Resource Economists*, 8:895–921.

Dijkstra, H. A. (2013). *Nonlinear Climate Dynamics*. Cambridge University Press.

Dupuis, P. and Ellis, R. S. R. S. (1997). *A weak convergence approach to the theory of large deviations*. Wiley series in probability and statistics. Probability and statistics. Wiley, New York ;.

Etminan, M., Myhre, G., Highwood, E. J., and Shine, K. P. (2016). Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, 43(24):12,614–12,623.

Flynn, C. M., Huusko, L., Modak, A., and Mauritsen, T. (2023). Strong aerosol cooling alone does not explain cold-biased mid-century temperatures in cmip6 models. *Atmospheric Chemistry and Physics*, 23(23):15121–15133.

Folini, D., Friedl, A., Kübler, F., and Scheidegger, S. (2024). The climate in climate economics*. *The Review of Economic Studies*, page rdae011.

Ghil, M. and Lucarini, V. (2020). The physics of climate variability and climate change. *Rev. Mod. Phys.*, 92:035002.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153.

Haensel, M. C., Drupp, M. A., Johansson, D. J. A., Nesje, F., Azar, C., Freeman, M. C., Groom, B., and Sterner, T. (2020). Climate economics support for the un climate targets. *Nature Climate Change*, 10:781–789.

Hansen, L. P. and Sargent, T. J. (2011). *Robustness*. Princeton University Press.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A., editors (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Intergovernmental Panel on Climate Change (IPCC).

Howard, P. H. and Sterner, T. (2017). Few and not so far between: A meta-analysis of climate damage estimates. *Environmental and Resource Economics*, 68:197–225.

Joos, F., Roth, R., Fuglestvedt, J. S., Peters, G. P., Enting, I. G., von Bloh, W., Brovkin, V., Burke, E. J., Eby, M., Edwards, N. R., Friedrich, T., Frölicher, T. L., Halloran, P. R., Holden, P. B., Jones, C., Kleinen, T., Mackenzie, F. T., Matsumoto, K., Meinshausen, M., Plattner, G.-K., Reisinger, A., Segschneider, J., Shaffer, G., Steinacher, M., Strassmann, K., Tanaka, K., Timmermann, A., and Weaver, A. J. (2013). Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis. *Atmospheric Chemistry and Physics*, 13:2793–2825.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2009). Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976.

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., Smith,

S. J., van den Berg, M., Velders, G. J. M., Vollmer, M. K., and Wang, R. H. J. (2020). The shared socio-economic pathway (ssp) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13:3571–3605.

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J. M., Vollmer, M. K., Wang, R. H. J., and Weiss, R. (2017). Historical greenhouse gas concentrations for climate modelling (cmip6). *Geoscientific Model Development*, 10:2057–2116.

Millner, A., Dietz, S., and Heal, G. (2012). Scientific ambiguity and climate policy. *Environmental and resource economics*, 55:21–46.

Nordhaus, W. (2008). *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press.

Rudik, I. (2020). Optimal climate policy when damages are unknown. *American Economic Journal: Economic Policy*, 12:340–373.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A. (2018). Fair v1.3: a simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*, 11(6):2273–2297.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R. (2020). Past warming trend constrains future warming in cmip6 models. *Science Advances*, 6(12):eaaz9549.

von der Heydt, A. S., Ashwin, P., Camp, C. D., Crucifix, M., Dijkstra, H. A., Ditlevsen, P., and Lenton, T. M. (2021). Quantification and interpretation of the climate variability record. *Global and Planetary Change*, 197:103399.

Xu, Z., Han, Y., Tam, C. Y., Yang, Z. L., and Fu, C. (2021). Bias-corrected cmip6 global dataset for dynamical downscaling of the historical and future climate (1979–2100). *Scientific Data*, 8(1):293.

Zhang, J., Furtado, K., Turnock, S. T., Mulcahy, J. P., Wilcox, L. J., Booth, B. B., Sexton, D., Wu, T., Zhang, F., and Liu, Q. (2021). The role of anthropogenic aerosols in the anomalous cooling from 1960 to 1990 in the cmip6 earth system models. *Atmospheric Chemistry and Physics*, 21(24):18609–18627.

Zhao, Y., Basu, A., Lontzek, T. S., and Schmedders, K. (2023). The social cost of carbon when we wish for full-path robustness. *Management Science*, 69:7585–7606. doi: 10.1287/mnsc.2023.4736.

Appendix A Derivations

A.1 Derivation of the minimizing weights under penalization

The relative entropy (or Kullback-Leibler divergence) for two probability measures π_1, π_2 over a finite space Θ is defined as

$$R(\pi_2|\pi_1) = \begin{cases} \sum_{\theta\in\Theta} \pi_2(\theta) \log\left(\frac{\pi_2(\theta)}{\pi_1(\theta)}\right) & \text{if } \pi_2 \ll \pi_1 \\ \infty & \text{else} \end{cases}$$

where we write $\pi_2 \ll \pi_1$ when π_2 is absolutely continuous with respect to π_1, meaning that $\pi_2(\theta) = 0$ whenever $\pi_1(\theta) = 0$. Two probability measures are said to be equivalent if they are absolutely continuous with respect to each other. Note that $R(\pi_2|\pi_1) \geq 0$, with equality if and only if $\pi_2 = \pi_1$.

THEOREM 1. *Let Π be the space of probability measures on the finite space Θ. Fix $\pi^* \in \Pi$ and $\kappa \in \mathbb{R}_{>0}$. For any function $f\colon \Theta \to \mathbb{R}$, the following variational formula holds:*

$$\min_{\pi\in\Pi} \sum_{\theta} \pi(\theta) f(\theta) + \kappa R(\pi|\pi^*) = -\kappa \log\left\{\sum_{\theta\in\Theta} \pi^*(\theta) \exp\left(-\frac{1}{\kappa} f(\theta)\right)\right\}.$$

The minimum is uniquely attained at π^o defined by

$$\pi^o(\theta) := \frac{\exp\left(-\kappa^{-1} f(\theta)\right) \pi^*(\theta)}{\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f(\theta')\right) \pi^*(\theta')}.$$

The proof is following closely section 1.4 in Dupuis and Ellis (1997), but adapted to a finite probability space.

PROOF. It suffices to consider the probability measures $\pi \ll \pi^*$ as they have finite relative entropy with respect to π^*. Define π^o as above. Note that π^o and π^* are equivalent, so that $\pi \ll \pi^*$ implied $\pi \ll \pi^o$. For any $\pi \ll \pi^*$,

$$\begin{aligned} &\sum_{\theta\in\Theta} \pi(\theta) f(\theta) + \kappa R(\pi|\pi^*) \\ &= \sum_{\theta\in\Theta} \pi(\theta) f(\theta) + \kappa \sum_{\theta\in\Theta} \pi(\theta) \log\left(\frac{\pi(\theta)}{\pi^o(\theta)}\right) + \kappa \sum_{\theta\in\Theta} \pi(\theta) \log\left(\frac{\pi^o(\theta)}{\pi^*(\theta)}\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\theta\in\Theta} \pi(\theta) f(\theta) + \kappa R(\pi|\pi^o) + \kappa \sum_{\theta\in\Theta} \pi(\theta) \log\left(\frac{\exp\left(-\kappa^{-1} f(\theta)\right)}{\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f(\theta')\right) \pi^*(\theta')}\right) \\
&= \sum_{\theta\in\Theta} \pi(\theta) f(\theta) + \kappa R(\pi|\pi^o) + \kappa \sum_{\theta\in\Theta} \pi(\theta) \log\left(\exp\left(-\kappa^{-1} f(\theta)\right)\right) \\
&\qquad - \kappa \sum_{\theta\in\Theta} \pi(\theta) \log\left(\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f(\theta')\right) \pi^*(\theta')\right) \\
&= \kappa R(\pi|\pi^o) - \kappa \log\left(\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f(\theta')\right) \pi^*(\theta')\right) \\
&\geq -\kappa \log\left(\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f(\theta')\right) \pi^*(\theta')\right)
\end{aligned}
$$

where the last step used that $R(\pi|\pi^o) \geq 0$. As equality holds if and only if $\pi = \pi^o$, π^o is the unique minimizing distribution. □

The objective function under misspecification aversion with a prior π^* follows from setting $f(\theta) := V(s,\ \mu,\ \theta)$ for any fixed policy pair $(s,\ \mu)$.

Note that the analogous derivation holds for time varying weights:

$$
\min_{(\pi_t)_t\in\Pi} \sum_{t=1}^{T} \left\{ \sum_{\theta\in\Theta} \pi_t(\theta) f_t(\theta) + \kappa R(\pi_t|\pi_t^*) \right\} = -\sum_{t=1}^{T} \kappa \log\left\{ \sum_{\theta\in\Theta} \pi_t^*(\theta) \exp\left(-\frac{1}{\kappa} f_t(\theta)\right) \right\}.
$$

The optimizing weights are then also time varying and given by

$$
\pi_t^o(\theta) := \frac{\exp\left(-\kappa^{-1} f_t(\theta)\right) \pi_t^*(\theta)}{\sum_{\theta'\in\Theta} \exp\left(-\kappa^{-1} f_t(\theta')\right) \pi_t^*(\theta')}.
$$

In our application,

$$
f_t(\theta) = \beta^t u(c_t|s,\mu,\theta).
$$

A.2 Social Cost of Carbon

The social cost of carbon is derived to be the shadow cost of an additional ton of emissions in terms of the numeraire good (consumption, or equivalently, capital). Let $\beta^t \xi_t$ be the multiplier on the emission accumulation constraint and $\beta^t \lambda_t$ be the multiplier on the consumption equation. First, note that if $\beta^t \tilde{\lambda}_t$ is the constraint on capital accumulation, then the first order condition for the savings rate shows that $\tilde{\lambda}_t = \lambda_t$ at the optimum[6]. The first order conditions on industrial emissions, overall emissions, greenhouse gas concentration and temperature then

[6]This holds whenever $s_t \in (0,1)$ which is the case in all of our runs

show that

$$\xi_t = \sum_{s=0} \sum_{j=0}^{s} \beta^{s+j} \frac{\partial M_{t+j}}{\partial E_t} \frac{\partial T_{t+s+j}}{\partial M_{t+j}} \frac{\partial Y_{t+s+j}}{\partial T_{t+s+j}} \lambda_{t+j+s}$$

The double sum stems from the path dependence in both the carbon cycle model and the temperature model. The dependence on carbon concentration can be made implicit to obtain the usual form

$$\frac{\xi_t}{\lambda_t} = \sum_{s=0} \beta^s \frac{u'(C_{t+s})}{u'(C_t)} \frac{\partial T_{t+s}}{\partial E_t} D'(T_{t+s}) Y^{ND}_{t+s},$$

where Y^{ND}_{t+s} is output before damages, but after mitigation expenditure. To obtain this shadow cost under uncertainty, note that the derivative of the objective, the expected value of the utility, with respect to either choice variable $x \in \{s\ , \mu\}$ is given by the expectation of the derivative. Under robustness concerns, the derivative takes the same form with an expectation formed under the tilted weights derived in the previous section A.1. This follows from

$$\begin{aligned} &\frac{\partial}{\partial x} \ -\kappa \log \left(\sum_{\theta} \pi(\theta) \exp \left(-\frac{1}{\kappa} V(s,\ \mu,\ \theta) \right) \right) \\ =&\frac{1}{\sum_{\theta} \pi(\theta) \exp\left(-\kappa^{-1} V(s,\ \mu,\ \theta)\right)} \sum_{\theta} \pi(\theta) \exp\left(-\kappa^{-1} V(s,\ \mu,\ \theta)\right) \frac{\partial V(s,\ \mu,\ \theta)}{\partial x}. \end{aligned}$$

The multipliers under uncertainty then satisfy

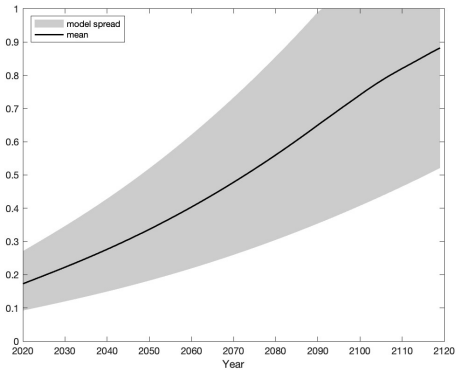
$$\frac{\xi_t}{\lambda_t} = \mathbb{E} \left\{ \sum_{s=0} \beta^s \frac{u'(C_{t+s})}{u'(C_t)} \frac{\partial T_{t+s}}{\partial E_t} D'(T_{t+s}) Y^{ND}_{t+s} \right\}$$

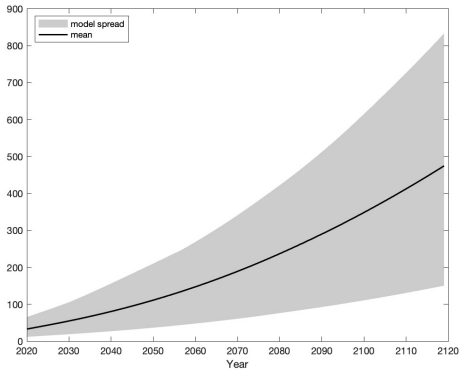
where $\mathbb{E}$ is the expectation reflecting the underlying subjective weights and robustness preferences.

Appendix B Additional Results

B.1 Lower damages

The damage function in Folini et al. (2024) as well as current versions of the DICE model uses a lower coefficient $\pi_2 = 0.0024$. Under the same value, the mean of our results are similar to the ones obtained by Folini et al. (2024). Mean optimal mitigation starts at around 20% in 2020, and the social cost of carbon at around 32$ per ton. No model suggests full mitigation before the year 2090. As pointed out by Haensel et al. (2020), the updated damage function is key to obtain IAM estimates that are consistent with UN climate targets.





Mitigation Rate Social Cost of carbon

Figure 10: Results under low damages (agnostic prior)